

# Modeling stochastic gene expression under repression

G.C.P. Innocentini, J.E.M. Hornos\*

Instituto de Física de São Carlos, Universidade de São Paulo,  
Caixa Postal 369, BR-13560-970 São Carlos, S.P., Brazil

## Abstract

Intrinsic transcriptional noise induced by operator fluctuations is investigated with a simple spin like stochastic model. The effects of transcriptional fluctuations in protein synthesis is probed by coupling transcription and translation by an amplificative interaction. In the presence of repression a new term contributes to the noise which depends on the rate of mRNA production. If the switching time is small compared with the mRNA life time the noise is also small. In general the dumping of protein production by a repressive agent occurs linearly but the fluctuations can show a maxima at intermediate repression. The discrepancy between the switching time, the mRNA degradation and protein degradation is crucial for the repressive control in translation without large fluctuations. The noise profiles obtained here are in quantitative agreement with recent experiments.

## 1 Introduction

The remarkable simplicity of the relation between genetic information and protein synthesis in the ribosomes, resulting in a symmetrical and almost universal genetic code, is the result of a complex process involving a large number of chemical reactions. The gene transcription is assisted by the enzymatic action of regulatory proteins which can enhance or repress the production of mRNA molecules. The amino acid assembling in the ribosome ends the cycle and proteins are produced and folded to a functional form. The pioneer work on lambda phage (1) have shown the crucial role of regulatory proteins in the control of such genetic networks (2) and intense experimental and theoretical investigation (3-13) have been dedicated to the understanding of this web of interactions.

A deterministic description of the whole process is hampered by the presence of a frequently small number of molecules in the cell and a stochastic approach is in general unavoidable (14-15). Several efforts have been directed to the numerical simulations of the genetic networks by considering the full set of chemical reactions (13-16). Alternatively one can write down a set of dynamical equations for protein concentrations followed by the inclusion of fluctuation by the Langevin mechanism (17-21).

A complete description of the cellular control mechanisms requires the understanding of the complex system of gene interactions and alternative efforts have been directed to the identification of elementary mechanisms to understand the random aspects of regulatory networks. A common feature of these models is the use of the rates of each chemical reaction to construct Markov process. A first class of models can be obtained by considering a single Poissonian model for transcription to be coupled to stochastic translation (23). The role of intrinsic noise can be analyzed in this framework by assuming an amplifier interaction designed to reproduced the observed protein bursts in translation (23-24). Alternatively regulation can be investigated by considering an unique effective stochastic model for the

---

<sup>2</sup>E-mail address: hornos@if.sc.usp.br

\*work supported by FAPESP and CNPq, Brazil

combined transcription-translation processes. In this context repression arises naturally by considering a spin like (27-28) or binary system (26) in the form of a switch. The advantage of spin like models is the possibility to use the well established many body techniques (29). Recently exact solutions for the Master equations of these system have been obtained in the case of an self interacting gene(30). In this article we combine the two strategies by adopting a two state birth and dead coupled model for the gene, in which the up state describes the free transcription of DNA while the down state represents repressed transcription, to be combined to a second Markov process for translation. In the absence of repression our calculations reproduces the model presented in reference (23). The regulation by an repressive environmental agent or protein concentration of another gene is fully considered in the model.

The transcription probabilities obey the equations

$$\frac{d\alpha_n}{dt} = k[\alpha_{n-1} - \alpha_n] + \rho[(n+1)\alpha_{n+1} - n\alpha_n] - h\alpha_n + f\beta_n \quad (1)$$

$$\frac{d\beta_n}{dt} = k\chi[\beta_{n-1} - \beta_n] + \rho[(n+1)\beta_{n+1} - n\beta_n] + h\alpha_n - f\beta_n \quad (2)$$

The stochastic variable  $n$  describes, here, the number of mRNA molecules in the cell. The unrepressed probability for RNA transcription is  $\alpha(t,n)$  and  $\beta(t,n)$  represented the partially dumped state of the switch. The RNA degradation rate in the cell is  $\rho$  with a life time  $T=1/\rho$ . The parameter  $k$  and  $\chi k$  are the free and repressed transcription rate, respectively. The parameter  $h$  and  $f$  controls the repressing efficiency of the switch. The capability of a protein to bind the repressor site is coded in the variable  $h$ . It is proportional to average repressing protein number. The unbinding rate reducing the repressing efficiency is  $f$ .

The effects of fluctuations on the mRNA population in the protein dynamics will be investigated in section VI by coupling this birth and dead process to a stochastic variable describing protein concentration. The generalization of this approach to a gene network requires the introduction of more random variables and the corresponding interactions.

This article is organized as its follows. In section II we solve the model by the introduction of generator functions. A single second order differential equation for the on state is obtained and reduced to the canonical form. In section III the physical content of the model is exhibit by the introduction of the relevant parameters describing the properties of the gene and environment. Simple expressions for the distributions, mean values and fluctuations are also presented. The induced noise is studied in section IV as a function of the switching time, transcription efficiency and repressing conditions. Microscopic functions are presented in section V. In section VI we couple the transcription to translation by the amplificative interaction of reference (23). The last section is dedicated to conclusions.

## 2 Master equations for repressed transcription

The discrete recursive equations implied in the master equations can be bypassed by the use of generating functions defined by (van Kampen, 1992)

$$\alpha(z, t) = \sum_{n=0}^{\infty} \alpha_n(t) z^n \quad \beta(z, t) = \sum_{n=0}^{\infty} \beta_n(t) z^n \quad (3)$$

The resulting master partial differential equations are

$$\frac{\partial \alpha(z, t)}{\partial t} = (z - 1)[k\alpha(z, t) - \rho \frac{\partial \alpha(z, t)}{\partial z}] - h\alpha(z, t) + f\beta(z, t) \quad (4)$$

$$\frac{\partial \beta(z, t)}{\partial t} = (z - 1)[\chi k\beta(z, t) - \rho \frac{\partial \beta(z, t)}{\partial z}] + h\alpha(z, t) - f\beta(z, t) \quad (5)$$

which must be solved under the condition  $\alpha(1, t) + \beta(1, t) = 1$  to ensure probability conservation. The moments can be easily obtained by the evaluation of this functions and its higher derivatives at  $z=1$ . The full information about the system requires the calculation of the distribution functions by the evaluation of derivatives at  $z=0$ :

$$\alpha_n(t) = \frac{1}{n!} \frac{d^n}{dz^n} \alpha(z, t) \quad \beta_n(t) = \frac{1}{n!} \frac{d^n}{dz^n} \beta(z, t) \quad (6)$$

This system is completely integrable due to the existence of Lie symmetries and the time dependence can also be obtained analytically. Here we focus our attention in the stationary properties of the system

$$(z - 1)(k\alpha_S + \rho \frac{d\alpha_S}{dz}) - h\alpha_S + f\beta_S = 0 \quad (7)$$

$$(z - 1)(\chi k\beta_S + \rho \frac{d\beta_S}{dz}) + h\alpha_S - f\beta_S = 0 \quad (8)$$

The coupled equation can be transformed in a single second order differential equation by replacing

$$\beta_S = \frac{1}{f} \left[ \rho(z - 1) \frac{d\alpha_S}{dz} - (zk - k - h)\alpha_S \right] \quad (9)$$

in equation (8). The resulting equation in the canonical form is

$$\frac{d^2 \alpha_S}{dz^2} + p \frac{d\alpha_S}{dz} + q\alpha_S \quad (10)$$

where the functions  $p$  and  $q$  are given by

$$p = \frac{(1 - z)(1 + \chi)k + \rho + h + f}{\rho(z - 1)} \quad (11)$$

$$q = \frac{k^2 \chi(z - 1) - k(\rho + f + \chi h)}{\rho^2(z - 1)} \quad (12)$$

The simple poles at  $z=1$  and the asymptotic irregularity suggests solutions in terms of the hypergeometric functions. In fact introducing the variable

$$\eta = \frac{k(\chi - 1)(z - 1)}{\rho} \quad (13)$$

and performing the transformation

$$\alpha_S = \exp\left(\frac{k(z-1)}{\rho}\right)\tilde{\alpha}(\eta) \quad (14)$$

we obtain the Kummer equation

$$\eta \frac{d^2 \tilde{\alpha}}{d\eta^2} + (b+1-\eta) \frac{d\tilde{\alpha}}{d\eta} - a\tilde{\alpha} = 0 \quad (15)$$

with parameters

$$a = \frac{h}{\rho} \quad (16)$$

$$b = \frac{h+f}{\rho} \quad (17)$$

The regular solution at  $z=1$  ( $\eta = 0$ ) can be written in terms of the Kummer M-functions

$$\alpha_S = \frac{b-a}{b} \exp\left(\frac{k(z-1)}{\rho}\right) M(a, b+1; \eta) \quad (18)$$

resulting in an analogous equation for  $\beta$

$$\beta_S = \frac{a}{b} \exp\left(\frac{k(z-1)}{\rho}\right) M(a+1, b+1; \eta) \quad (19)$$

A simple formula can also be obtained for the total probabilities

$$\phi_S = \exp\left(\frac{k(z-1)}{\rho}\right) M(a, b; \eta) \quad (20)$$

The advantage of the use of generating functions is the analytical elimination of the singular solution (a U-Kummer function). In numerical calculations the presence of this component is the source of numerical instabilities in the algorithm

### 3 The biological content of the model

The relevant information about the properties of the switch in the proteomic environment is specified by two time scales

$$T = \frac{1}{\rho} \quad \tau = \frac{1}{h+f} \quad (21)$$

The mean life of the mRNA molecules in the cells is coded in  $T$ , it is typically of order of minutes. The meaning of the switching time  $\tau$  is related to the time evolution of the stochastic dynamics process. In fact, a simple equations for the probabilities  $p_\alpha(t)$  and  $p_\beta(t)$  to find the gene free or repressed by a protein in the operator site is obtained by setting  $z=1$  in equations (1) and (2):

$$\frac{dp_\alpha(t)}{dt} = -hp_\alpha(t) + fp_\beta(t) \quad (22)$$

$$\frac{dp_\beta(t)}{dt} = +hp_\beta(t) - fp_\beta(t) \quad (23)$$

summing Eq.2 over n with the use of probability conservation. The solution is:

$$p_\beta(t) = \frac{h}{h+f} + e^{-(h+f)t/\rho} (p_\beta(0) - \frac{h}{h+f}) \quad (24)$$

The switching time is a intrinsic property of the gene. It doesn't depends on the birth and dead constants,  $k$  and  $\rho$ . The adimensional parameter

$$\epsilon = \frac{h+f}{\rho} \quad (25)$$

contains the information about the flexibility of the gene switch. High values of  $\epsilon$  corresponds to fast switches reaching the equilibrium configuration before the gene transcription is stationary. As we show above the average number of mRNA copies doesn't depends on the switching time but the noise is strongly affected by this parameter.

The influence of the environment on the gene transcription is measured by the asymptotic probability to find the gene repressed

$$p_\beta = \frac{h}{h+f} \quad p_\alpha = \frac{f}{h+f} \quad (26)$$

In the limit  $p_\beta = 0$  we recover the model presented in (23) with only one transcription mode and a Poissonian distribution where  $N = \frac{k}{\rho}$  is the average number of mRNA copies. Highly repressed genes belongs to the  $p_\beta = 1$  sector where the mRNA transcription is suppressed or occurs in the repressed mode ( $\chi \neq 0$ ). The parameter  $\chi$  takes in account the possibility of a small rate of transcription in the (partially) repressed mode. In summary,  $N$  measures the efficiency of the transcription. A gene with a large value of  $N$  produce mRNA in abundance, in the absence of repression, but also can be damped by regulatory proteins or an external factor to  $\chi N$  particles in average. Inefficient transcriptions as occurs in certain genes of E.Coli, for example, corresponds to  $N$  around one or two, while in copious transcription  $N$  increases an order of magnitude. The effects of the protein concentration or another repressing factor is controlled by  $p_\beta$  which can be phenomenologically related to protein concentration by a Hill function, for example. The relative switch frequency scales the capability of the gene to flip from an arbitrary condition to it's asymptotic value. The relevant features of the probabilities can be obtained by inspecting the generator functions written in terms of these parameters:

$$\alpha_S = p_\alpha \exp(N(z-1)) M(\epsilon p_\beta, \epsilon + 1; N(z-1)(\chi-1)) \quad (27)$$

for the up state

$$\beta_S = p_\beta \exp(N(z-1)) M(\epsilon p_\beta + 1, \epsilon + 1; N(z-1)(\chi-1)) \quad (28)$$

for the down state

$$\phi_S = \exp(N(z-1)) M(\epsilon p_\beta, \epsilon; N(z-1)(\chi-1)) \quad (29)$$

for the total probability

In the absence of a repressive agent  $p_\beta = 0$  and  $p_\alpha = 1$ , the Kummer function reduces to one and we have a Poissonian process producing  $N$  mRNA copies in average. In general the generator function is a power series on  $N$ , modulated by  $\chi$  with coefficients depending of the repression parameter and switching time via a Pochhammer function. If the gene interact with a stochastic factor, as in the case of fluctuating proteins due to another gene, the parameters  $N$ ,  $p_\beta$  and eventually  $\epsilon$  should be considered random functions of the corresponding proteomic field. In the model we have two different sources for hampering or enhancing transcription. The transcription efficiency caused by the availability of polymerases affects the free mean number  $N$  while the density of repressive proteins modifies  $p_\beta$ .

If we introduce a power series in the above equation we obtain recursion relations for a second order Markov process for the total probability. The probability is obtained taking successive derivatives of the generator function by the use of the Leibnitz rule and the formulas for the derivatives of Kummer functions:

$$\alpha_n = p_\alpha (N)^n \frac{\exp(-N)}{n!} \sum_{s=0}^n \binom{n}{s} (\chi-1)^s \frac{(p_\beta \epsilon)_s}{(\epsilon+1)_s} M(p_\beta \epsilon + s, \epsilon + 1 + s; N(1-\chi)) \quad (30)$$

$$\beta_n = p_\beta (N)^n \frac{\exp(-N)}{n!} \sum_{s=0}^n \binom{n}{s} (\chi-1)^s \frac{(p_\beta \epsilon + 1)_s}{(\epsilon+1)_s} M(p_\beta \epsilon + 1 + s, \epsilon + 1 + s; N(1-\chi)) \quad (31)$$

$$\phi_n = (N)^n \frac{\exp(-N)}{n!} \sum_{s=0}^n \binom{n}{s} (\chi-1)^s \frac{(p_\beta \epsilon)_s}{(\epsilon)_s} M(p_\beta \epsilon + s, \epsilon + s; N(1-\chi)) \quad (32)$$

The Poisson distribution appears with an envelope function composed by a superposition of distributions implementing the repression. In figure (1) we show a typical distribution for the efficient transcription of gene producing  $N=40$  mRNA copies. The repressed mode produces 20 of the free mode ( $\chi = 1/5$ ). The gene is slow  $\epsilon = 1/2$ . The  $\alpha$  and  $\beta$  components are shown together with the sum in the figure. The mean peak corresponds to the free production and the secondary to the repressed transcription.

## 4 Mean values and noise

The closed form for the generator functions and the friendly properties of the Kummer functions allows one to obtain simple formulas for all the distribution moments evaluating the derivatives at the point  $z=1$ . The mRNA mean value produced in both models in the cell is

$$\langle n \rangle = N(p_\alpha + \chi p_\beta) \quad (33)$$

It is linear in the gene efficiency  $N$  and in the repression parameter  $p_\beta$ . The switch life time do not appear in the mean value. If we assume that the probability to find the gene in the on mode without a binding repressing protein is related to the concentration by a Hill function

$$p_\alpha = \frac{1}{1 + (\langle m \rangle / K)^l} \quad (34)$$

where  $K$  is the threshold concentration of the repressor agent and  $l$  is steepness factor, we obtain the curves shown in Fig.2. The higher curve corresponds to a gene that can be transcribed poorly even in the presence of the repressor agent ( $\chi = 1/5$ ). The free transcription is not affected by low concentrations but will decay abruptly after the threshold concentration when the mean values reaches it's repressed value. The second curve describes a  $\chi = 0$  switch and after threshold there is no transcription in the cell.

The possibility of repressed production causing the appearance of two peaks statistical distributions requires the use of the moments of the on and off components. In this case the total mean value and also the total fluctuation will have a limited meaning. However simple formulas can also be obtained for the partial mean values:

$$\langle n \rangle_\alpha = \langle n \rangle + \frac{N(1-\chi)p_\beta}{1+\epsilon} \quad (35)$$

$$\langle n \rangle_\beta = \langle n \rangle - \frac{N(1-\chi)p_\alpha}{1+\epsilon} \quad (36)$$

The effects of the transcription efficiency and the switching frequency in the noise can be evaluated from the standard deviation

$$\langle n^2 \rangle - \langle n \rangle^2 = \langle n \rangle + \frac{N^2(1-\chi)^2 p_\alpha p_\beta}{1+\epsilon} \quad (37)$$

In both limits  $p_\beta = 0$ , absence of repression, and  $p_\beta = 1$ , total repression, the stochastic process is Poissonian. The square root deviation as a function of the repressing parameter  $p_\beta$  is shown in Fig.3 for the copious transcription of a gene. The curves correspond to several values of the switching frequency. For slow genes (small  $\epsilon$ ) the noise increases with the repressing parameter until maximal fluctuations are reached for

$$(p_\beta)_{max} = \frac{1}{2} - \frac{1}{2} \frac{1+\epsilon}{N(1-\chi)} \quad (38)$$

After this limit the fluctuations will be smaller until a minimal value for total repression. In this case, to reach the stationary configuration, the gene will intercalate on and off configurations many time causing strong noisy in the copious transcription. We are describing the noise in terms of the mean deviation instead of the Fano factor that is exhaustive used in physics. The two alternatives are of course equivalent but the Fano factor measures the deviation from a Poisson distribution and don't have the intuitive appeal of the distribution width. For large values of  $\epsilon$  corresponding to fast switches the parabolic fluctuations will not reach the maxima and decreases monotonically. The repression drops the mean value of the mRNA transcription and the fluctuations simultaneously. Inspecting equation (37) we see the presence of a factor  $\epsilon + 1$  in the denominator responsible for the noise attenuation. The disparity between the mRNA life time in the cell and the binding-unbinding characteristic time leading to large values of  $\epsilon$  ensures low fluctuations and silent transcription. In Fig.4 the fluctuations of a poorly transcribed gene are shown. The enhancement of the noise is limited by the presence of few mRNA copies in the cell and suppressed totally for fast switches. The critical switch time is

$$(\epsilon)_{crit} = N(1-\chi) - 1 \quad (39)$$

We should keep in mind that the abundant transcription of a gene requires the enzymatic action to eliminate the unused messengers expending energy.

## 5 Probabilities

The description of the random gene dynamics in terms of the two first moments are satisfactory for local statistical distributions which is the most common case. The fluctuations will describe properly the departure from the deterministic or macroscopic behaviour. However the possibility of repressed production or the appearance of strongly non local probabilities requires the full knowledge of the probability distribution. The solubility of the model allow us to investigate the behaviour of the probabilities as a function of the biophysical parameters of the model.

The probabilities as a function of mRNA population for abundant transcription of a gene is shown in Fig.3. We assume that without repression  $N=40$  mRNA copies would be produced in the stationary state. A small repressed production  $\chi = 1/5$  is also allowed. In Fig.3A the behaviour of a slow switch,  $\epsilon = 1/2$ , is investigated as a function of the repressor activity  $p_\beta$ . In the absence of repression we see a Poissonian peak around  $n=40$ . The presence of moderated repression,  $p_\beta = 0.25$  has three main effects: reduces the Poissonian maximum, displaced toward lower population and creates a secondary peak at lower population. The switch is been turned off by the action of the repressing factor. The situation is already reversed by intermediate repression, the secondary maximum now is dominant indicating that the transcription of the gene occurs only in the off, or partially repressed state. Under strong repression  $p_\beta = 0.75$  the Poissonian peak disappears totally and finally under a full repression the gene is off been transcribed secondarily at low rates. Increasing the switching time will favour the delocation of the probability as shown in Fig.3B. The free and the totally repressed curves are the same as in Fig.3A but the intermediate curve  $p_\beta = 0.5$  is widespread over the population. The situation is even more dramatic if the gene is faster as is shown in Fig.3C. A plateau with equally probable population appears under intermediate repression. The mean value and fluctuation have a limited meaning in this case. In Fig.3C we can see the disappearance of the double peaks under any repressive condition. Finally in Fig.3E and Fig.3F the probabilities recovers the Poissonian shape and moves adiabatically under the repression action. In Fig.3E as a solitary wave keeping the shape undisturbed and in the case of very fast switches (Fig.3F) decreasing the mean deviation until total repression. The turn off of a gene by environmental repression is shown in Fig.4A for the case of efficient transcription. We choose  $\chi = 0$  therefore there is no residual transcription in the off mode. The first curve, incomplete in the plot, represents the Poissonian unrepressed distribution, and the other shows the evolution of the probability for growing repression. A kink is formed for intermediate repression ( $p_\beta = 0$  centered around  $30 < n < 40$  separating the equally distributed probability region from the null region. The peak concentrate at  $n=0$  for  $p_\beta = 1$ ) shows the total repression of the transcription. The inefficient transcription is shown in figures 4B and 4C. The last corresponds to high values of  $\epsilon$  and shows the typical Poissonian behaviour. However in figure 4B we can see the delocation of the statistical distribution typical of slow switches. In the set of figures, 5A, 5B, 5C, we show the tridimensional plot of the probabilities as a function of  $\epsilon$  for null, intermediate and total repression. At intermediate repression the  $\epsilon$  variation transform the double peak distribution at low  $\epsilon$  in a non-local function until the Poisson like behaviour reappears for fast switches. The critical value of  $\epsilon$  for which the fluctuation decreases without an enhancement of fluctuation is given by

In Fig.5 the probability surfaces are show as a function of the population and switching time for unrepressed, intermediate and fully represses systems. The switching time can change considerably the distribution pattern at intermediate repression.

## 6 Induced noise in protein production

The effects of the transcription fluctuations on the protein concentration can be investigate allowing our probabilities to depend on a new stochastic variable  $m$  for the protein concentration. This will be refrased in the language of generator functions by the appearance of a new continuous variable  $y$



and we will deal with a time dependent partial differential equation in both variables. Among a large variety of probability conserving coupled partial equations the simplest is given by:

$$\frac{\partial \alpha}{\partial \tau} = (z - 1)[N\alpha - \frac{\partial \alpha}{\partial z}] - p_\beta \epsilon \alpha + p_\alpha \epsilon \beta + (y - 1)[\sigma z \frac{\partial \alpha}{\partial z} - \eta \frac{\partial \alpha}{\partial y}] \quad (40)$$

$$\frac{\partial \beta}{\partial \tau} = (z - 1)[\chi N \beta - \frac{\partial \beta}{\partial z}] + p_\beta \epsilon \alpha - p_\alpha \epsilon \beta + (y - 1)[\sigma z \frac{\partial \beta}{\partial z} - \eta \frac{\partial \beta}{\partial y}] \quad (41)$$

The parameters  $\sigma$  and  $\eta$  are equal to  $k_P/\rho$  and  $\rho_P/\rho$ , respectively, where  $k_P$  and  $\rho_P$  are the birth and dead rates of the protein production effective reaction. This amplifier coupling reproduces the bursts in protein production that are observed experimentally. For  $p_\beta = 0$  the transcription is Poissonian and the usual results are obtained. However in the general case both effects, the induced noise caused by the coupling and the repressive noise are present. Although the full distributions are hard to obtain, the moments can be recursively obtained by the tradition techniques (Kampen, 1992). The mean value for protein concentration is

$$\langle m \rangle = \frac{\sigma N}{\eta} (\chi p_\beta + p_\alpha) \quad (42)$$

The adimensional parameter  $\eta$  is small, typically around 1/30 or less due to the discrepancy between the time scales for transcription and translation. The mRNA free mean number  $N$  is amplified by a factor  $\sigma/\eta$ . The fluctuations are

$$\langle m^2 \rangle - \langle m \rangle^2 = \langle m \rangle + \frac{\sigma^2 N}{\eta(\eta + 1)} \left( p_\alpha + \chi p_\beta + \frac{N p_\alpha p_\beta (\eta + \epsilon + 1) ((\chi - 1)^2)}{(\eta + \epsilon)(\epsilon + 1)} \right) \quad (43)$$

The terms on parenthesis are caused by the environmental repression. The relative frequencies  $\epsilon$  and  $\eta$  define three time scales for the phenomena. The switching time associated to the binding-unbinding of proteins to the operator site is the smaller time in the problem. A second scale, one order of magnitude higher, is defined by the mRNA degradation rate. Finally the protein degradation rate is the slower reaction time. Ultimately the slow down of the noise in the global process is ensured by the discrepancy of these scales and guarantee the precision in cellular reproduction. In Fig. 6 we show the standard deviation in three cases with the parameters adjusted to give the same protein mean concentration in the cell, ( $\langle m \rangle = 3200$ ) in order to compare the two possible strategies. The monotonically decreasing function shows the behaviour of the fluctuation for inefficient transcription ( $N=4$ ). The quadratic dependence of the variance on the protein parameter  $\sigma$  is responsible for the large deviation in the absence of repression. The second curve correspond to  $N$  around 6 and  $\epsilon=5$ . The crossing point around  $p_\beta \approx 0.5$  correspond to states with the same mean number, fluctuation and repression but with different switch velocities and production rates. In the third curve fluctuations are enhanced parabolic by low repression until a maxima is obtained. Beyond this value the deviation is attenuated until the system is off. For intermediate values of  $p_\beta$  both process are equivalent resulting in comparable mean values and noise. Finally we consider a gene with an increasing activity by an inductor A concrete example is a tetracycline repressor under the control of IPTG in E. Coli. This gene have been used as a part of a synthetic gene cascade in recent experiments. Instead of the mean square root deviation of fano factor we display the noise using the quantity

$$\Delta^2 = \frac{\langle m^2 \rangle - \langle m \rangle^2}{\langle m \rangle^2} \quad (44)$$

We relate the model parameter  $p_\alpha$  with the repressor concentration by a Hill function

$$p_{\alpha} = \frac{(c/k)^{\theta}}{1 + (c/k)^{\theta}} \quad (45)$$

where  $c$  is the concentration and  $k$  a parameter. The protein mean value and the corresponding noise are shown in figure 7.A and 7.B and have been obtained replacing  $p_{\alpha}$  from equation (45) in equations (42) and (43), respectively. It is remarkable the agreement between the concentration and the noise profiles with the experimental data of reference (25). In the present case we have single gene, global fluctuation have been discarded and  $p_{\alpha}$  depends only on the mean concentration of the inductor. In figure 8.A and 8.B we show the opposite situation in which the concentration decrease and the noise increases. We can see in the figure a maxima in the fluctuation for intermediate concentrations. Again the profile is similar to experimental results even under the mentioned differences and restrictions.

## 7 Conclusions

The hampering of noise in genetic networks by the action of regulatory proteins is supported by some experiments and also by the common intuition that repression slow down fluctuations. However this is not a mandatory rule. The presence of fluctuations is governed by the transcription efficiency of the gene, by the repressor concentration but also by the switching time describing the gene agility to interchange from a repressed state to a free mode. These components combined in a simple stochastic model for repressed transcriptions allow us to establish conditions for the occurrence of noise in a elementary network component.

The transcription of a gene at slow rates is silent even if the gene switch turn off slowly. The repression agent decreases the fluctuations in accord to the concentration of the controlling factor. In the common case of fast switches the microscopic distributions are Poisson-like and the flip-flop of the gene configurations induced by repression will cause small corrections.

In the opposite case when the transcription is copious the mRNA population fluctuates in the cells strongly in the case of slow switches. The noise is smaller than in the unrepressed case only for ultra fast switches revealing the fundamental role of the time switch in the dynamics of the noise.

The microscopic probabilities obtained due to the integrability of the coupled Markovian process shows non-local, broad bands population distributions under highly noisy conditions. The possibility of partial repressed transcription allowed in the model causes the appearance of double peaks in the probabilities due to the repressed and free transcription.

The two fundamental mechanisms which causes the decreasing in transcription rates, the shortage of polymerases and the presence of a repressor agent affects the fluctuations differently. In the first case the average mRNA population decreases homogeneously with the gene efficiency parameter. In the second case we see the parabolic dependence with the repression parameter.

The coupling of the transcription to translation by the amplifying mechanism reproduces the occurrence of protein bursts under repression with a rich noise structure.

Our model is minimal in the sense that: (i) It is a simple and soluble second order stochastic model for repressed transcription. (ii) The introduction of two states for the gene contemplating the repressed transcription introduces the two time scales involved in the process. (iii) The effects on translation proceeds by minimally coupling the transcription to protein production. (iv) Gene interactions and auto regulation can be implemented by allowing fluctuations in the model parameters.

The evolutionary requirements of noise to enforce differentiation, prompts the need of experiments, beyond the poorly transcribed genes, as maltT of E.coli, designed to probe the noise in copious transcription.

The comparison of our model with experiments shows a good agreement even considering the the limitations of our single gene treatment to explain the observed cascade effects under global noise.

Acknowledgment: J.E.M. Hornos wishes to thanks to J.N. Onuchic for the introduction into the field, P. Wolynes for helpful discussions and W. Arber for the memories on lambda phage experiments.

## References

- [1] Ptashne, M (1992) *A Genetic Switch: Phage  $\lambda$  and Higher Organisms*. Cambridge, MA:Cell Press/Blackwell.
- [2] Monod, J and Jacob, F (1962) *Cell. Regul. Mech* **june**, 4-12.
- [3] Kennell, D. and Tiezman, H. (1977) *J. Mol. Biol.* **114**, 1-21.
- [4] Gillespie, D.T. (1977) *J. Phys. Chem.* **81**, 2340-2361.
- [5] Berg, O.G. (1978) *J. Theor. Biol.* **71**, 587-603.
- [6] Ackers, G.K., Johnson, S.D. and Shea, M.A. (1982)*Proc. Natl. Acad. Sci. USA* **79**, 1129-1133.
- [7] Ko, M.S.H. (1991) *J. Theor. Biol.* **153**, 181-192.
- [8] van de Putte, P. and Goosen, N. (1992) *Trends Genet.* **8**, 457-462.
- [9] Siegele, D.A. and Hu, J.C. (1997) *Proc. Natl. Acad. Sci.USA* **94**, 8168-8172.
- [10] Cook, D.L., Gerber, A.N. and Tapscott, S.J. (1998) *Proc. Natl. Acad. Sci.* **95**, 15641-15646.
- [11] Bhalla, U.S. and Iyengar, R. (1999) *Science* **283**, 381-387.
- [12] von Dassow, G., Meir, E., Munro, E.M. and Odell, G.M. (2000) *Nature (London)* **406**, 188-192.
- [13] Becskei, A. and Serrano,L. (2000) *Nature (London)* **405**, 590-593.
- [14] McAdams, H.H. and Arkin, A. (1999) *Trends Genet.* **15**, 65-69.
- [15] McAdams, H.H., and Arkin, A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 814-819.
- [16] McAdams, H.H., and Arkin, A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 814-819.
- [17] Gardner, T.S., Cantor, C.R. and Collins, J.J. (2000) *Nature (London)* **403**, 339-342.
- [18] Hasty, J., Pradines, J., Dolnik, M. and Collins, J.J. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2075-2080.
- [19] Blake, W.J., Kaern, M., Cantor, C.R. and Collins, J.J. (2003) *Nature* **422**, 633-637.
- [20] Paulsson, J., Berg, O.G. and Ehrenberg, M. (2000)*Proc. Natl. Acad. Sci. USA* **97**, 7148-7153.
- [21] Elowitz, M.B. and Leibler, S. (2000)*Nature (London)* **403**, 335-338.
- [22] Elowitz, B.E., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) *Science* **297**, 1183-1186
- [23] Thatai, M. and van Oudenaarden, A. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 8614-8619.
- [24] Ozbudak, E.M., et al., (2002) *Nature Genet.* **31**, 69-73.
- [25] Pedraza, J.M. and van Oudenaarden, A. (2005) *Science* **307**, 1965-1969.

- [26] Kepler, T.B. and Elston, T.C. (2001) *Biophys. J.* **81**, 3116-3136.
- [27] Sasai,M. and Wolynes,P.G. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 2374-2379
- [28] Ralf Metzler and Peter G. Wolynes (2002) *Chemical Physics* **284**, 469-479
- [29] Walzack, A.M., Sasai, M. and Wolynes, P.G.(2005) *Biophys. J.* **88**, 828-850
- [30] Hornos, J.E.M., Schultz, D., Innocentini, G.C.P., Wang, J., Walzack, A.M., Onuchic, J.N. and Wolynes, P.G.(2005) *Phys. Rev. E* **72**, 05197-1 05197-5

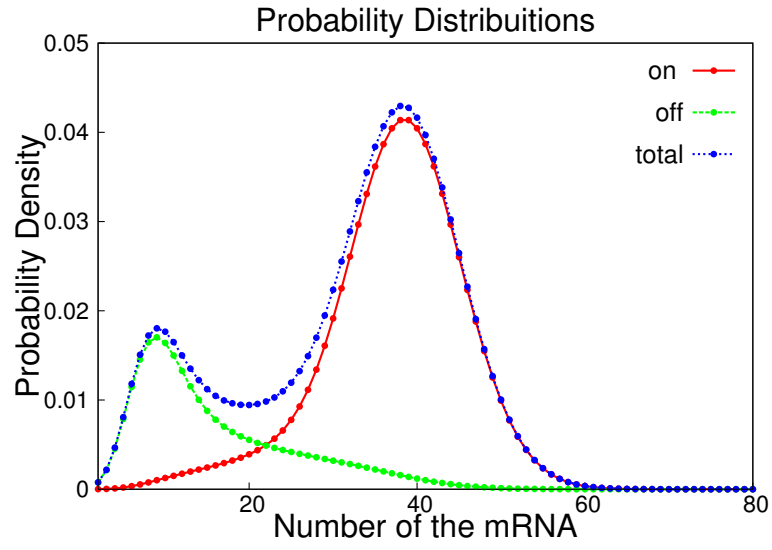


Figure 1: Probabilities for the on, off states are show together with the total probability. The parameters are  $N=40$ ,  $\chi=1/5$ ,  $\epsilon = 1/2$  and  $p_{\beta} = 0.20$

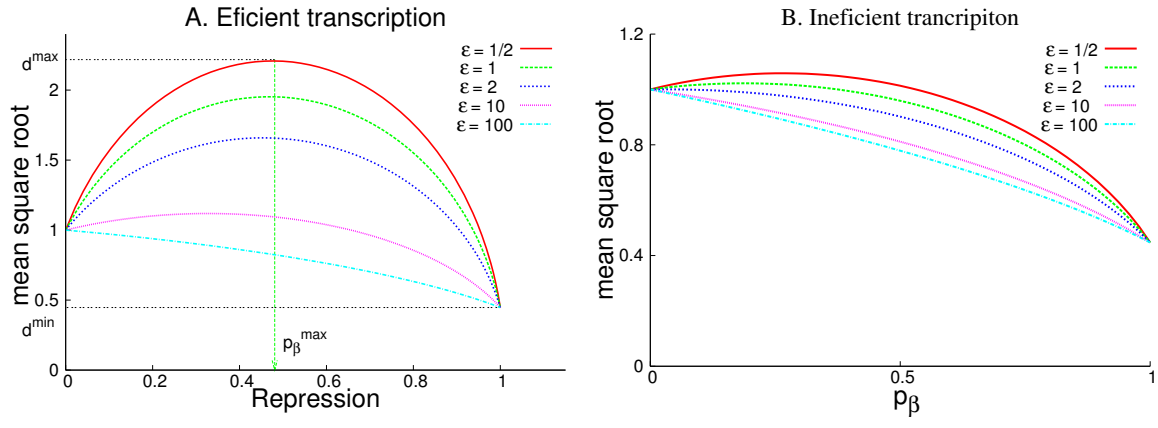


Figure 2: Figures 2A and 2B, standard deviation for high and low transcription with parameter  $N=40$  and 4 respectively. Repressed production  $\chi = 1/5$

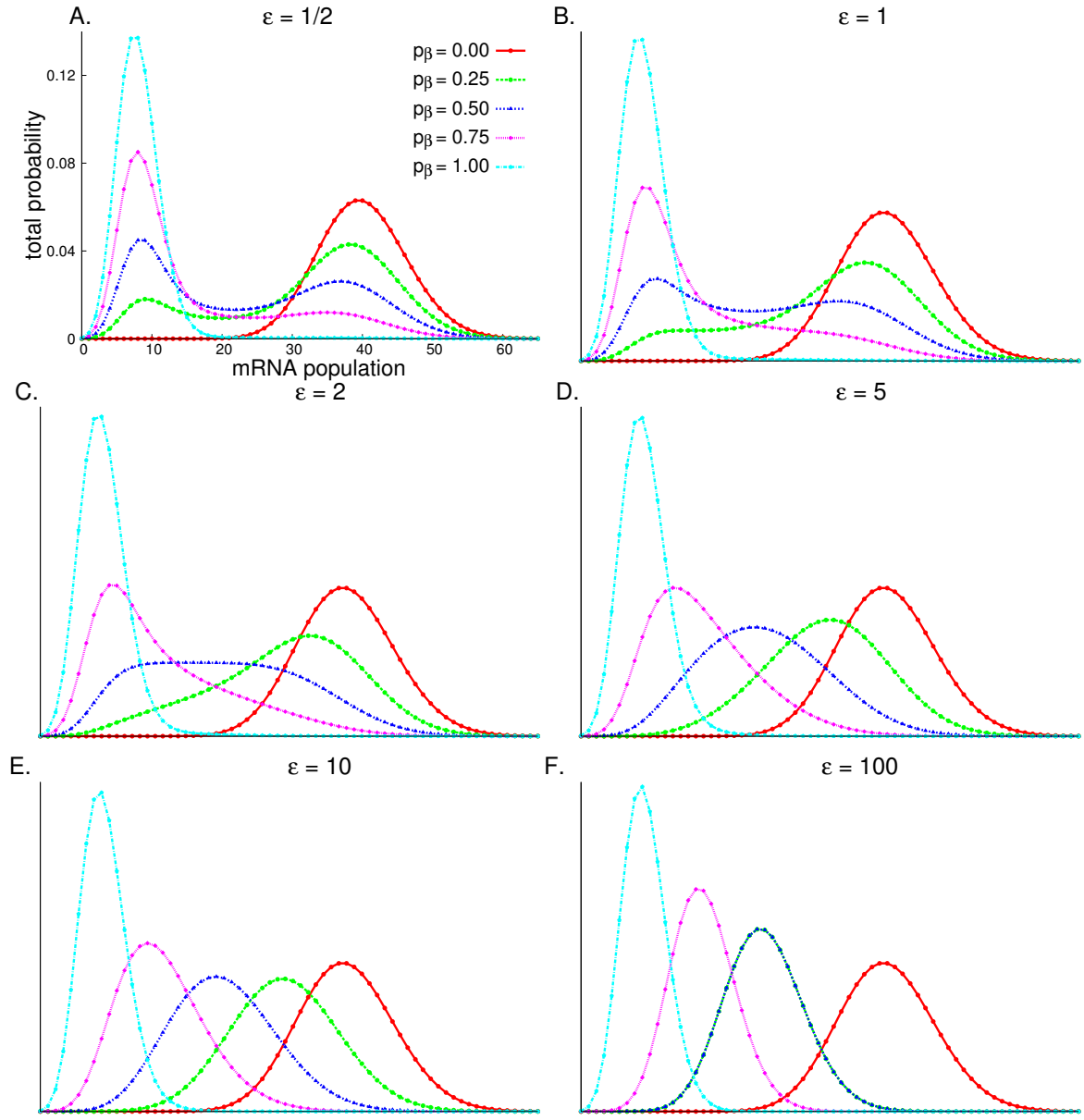


Figure 3: Efficient transcription,  $N=40$ .  $\chi$  is chosen  $1/5$ . The values of  $\epsilon$  and  $p_\beta$  are shown in the figure.

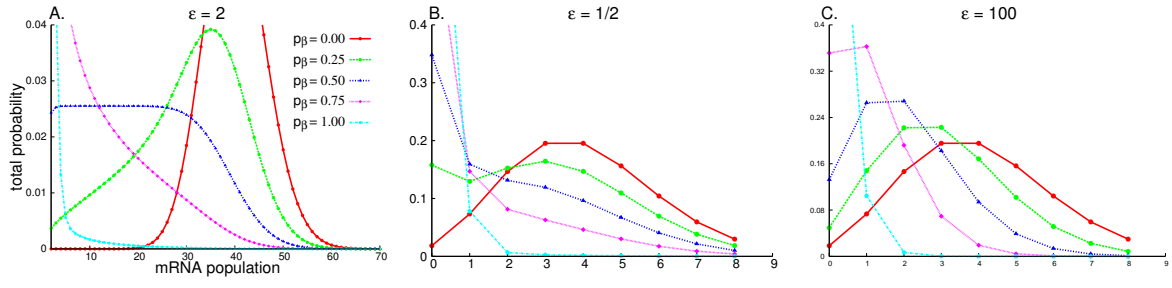


Figure 4: In Fig.4A  $\epsilon=2$  and  $N=40$ . Figures 4B and 4C,  $N=4$  and  $\epsilon=1/2, 100$ , respectively.  $\chi=1/50$  in all cases



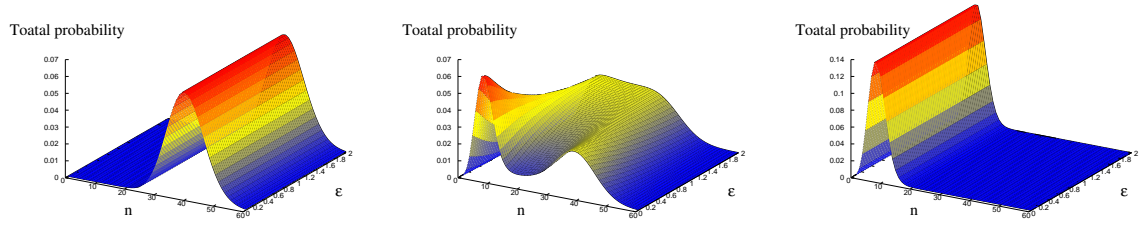


Figure 5: Surfaces for  $N=40$  and  $\chi=1/5$ . In, A, B and C  $p_\beta=0, 1/2$  and  $1$ , respectively

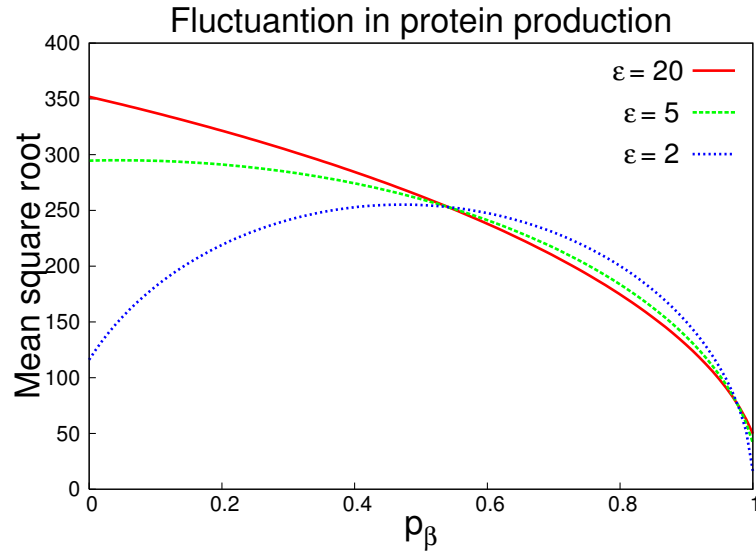


Figure 6: Standard deviation in protein concentration. For all curves  $\eta = 1/20$  and  $\chi = 1/50$ . The curve with  $\epsilon = 20$  the parameters are:  $K_P = 39.60$  and  $K_R = 4.04$ . The curve with  $\epsilon = 5$ ,  $K_P = 27.44$  and  $K_R = 5.83$  and the curve with  $\epsilon = 2$ ,  $K_P = 3.37$  and  $K_R = 47.48$ .

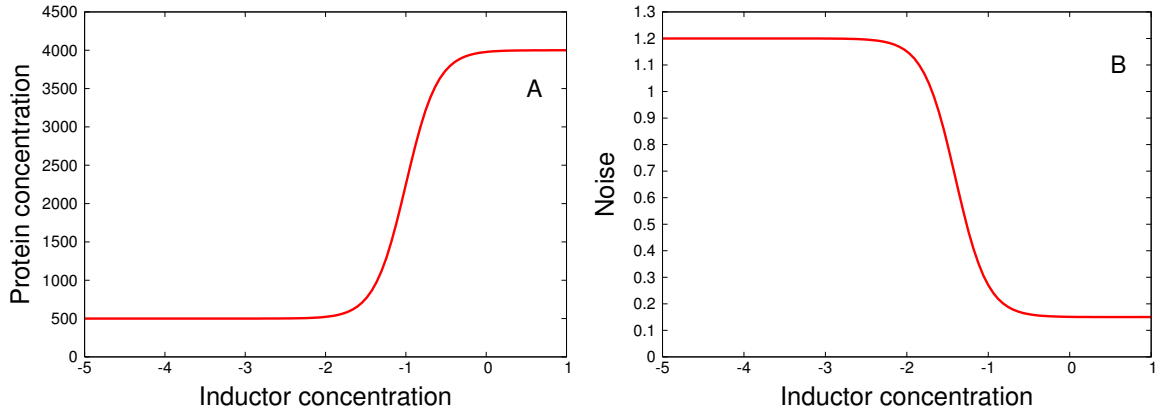


Figure 7: In figure 7.A and 7.B the parameters are:  $\sigma = 730$ ,  $N = 1.20$ ,  $\eta = 0.218$ ,  $\chi = 1/8$ ,  $\epsilon = 30$ ,  $k = 1/10$  and  $\theta = 2.2$ . In figure 7.B noise gene by equation (44) as a function of the inductor concentration.

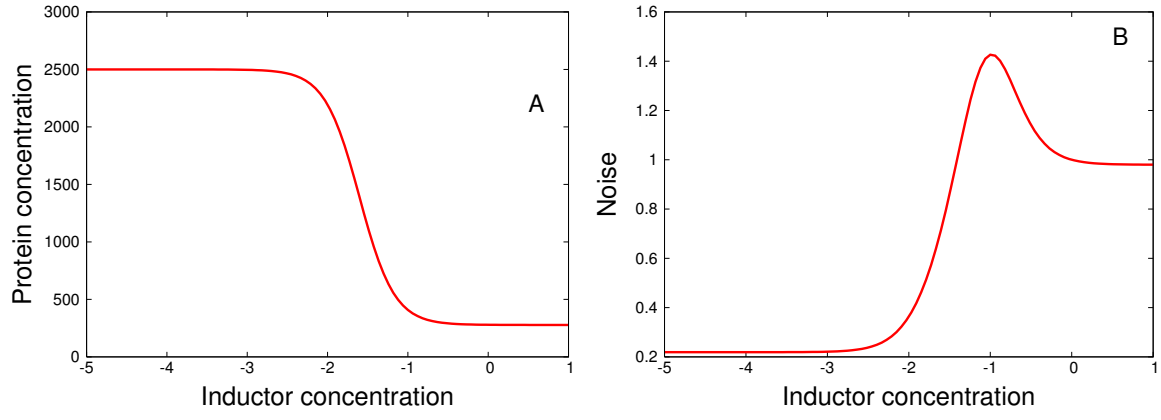


Figure 8: In both figures the parameters are:  $\sigma = 125$ ,  $N = 1$ ,  $\eta = 1/20$ ,  $\chi = 1/20$ ,  $\epsilon = 1/10$ ,  $k = 1/40$  and  $\theta = 2$ . In figure 8.B noise gene by equation (44) as a function of the inductor concentration.